Datenbanken
Wintersemester 2020/21
Prof. Dr. W. May

2. Übungsblatt: Algebra

Besprechung voraussichtlich am 9./10./16./17.12.2020

Aufgabe 3 (Relationale Anfragen an Mondial: Schweizer Sprachen) Geben Sie Ausdrücke der relationalen Algebra für die folgenden Anfragen an die Mondial-Datenbank an:

a) Alle Landescodes von Ländern, in denen eine Sprache gesprochen wird, die auch in der Schweiz gesprochen wird.

b) Alle Landescodes von Ländern, in denen ausschliesslich Sprachen gesprochen werden, die in der Schweiz nicht gesprochen werden.

c) Alle Landescodes von Ländern, in denen nur Sprachen gesprochen werden, die auch in der Schweiz gesprochen werden.

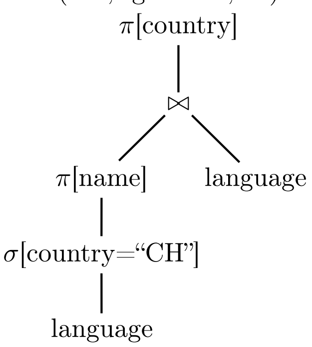
d) Alle Landescodes von Ländern, in denen alle Sprachen gesprochen werden, die in der Schweiz gesprochen werden.

Für spätere Übungsblätter:

- Geben Sie dieselben Anfragen in SQL an.

---

a) Join zur Formulierung einer "Auswahlbedingung". Verwende Tabelle language(country,name,percent), z.B. (CH, "german", 65).

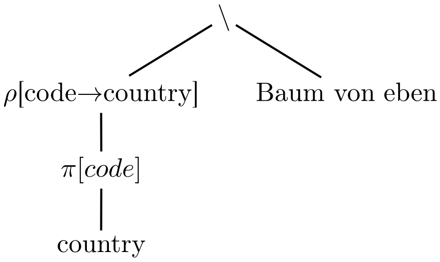


Der linke Ast ergibt die Tabelle

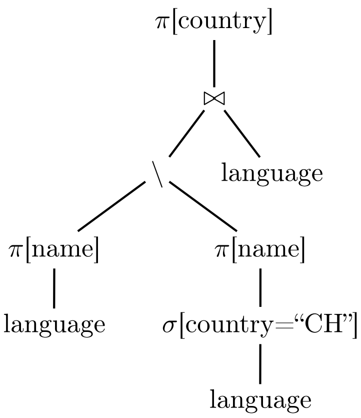
language
name
german
french
italian
romansch

Diese wird mit language gejoint (Join-Attribut "name"), womit genau die Einträge aus language übrigbleiben, die eine der aufgeführten Sprachen enthalten.

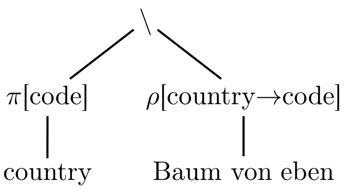
b) Genau alle Länder, die eben nicht aufgezählt wurden:



c) Löst man in mehreren Schritten. Erst alle Sprachen bestimmen, die nicht in CH gesprochen werden. Das Ergebnis wird mit language gejoint, und übrig bleiben die Einträge, die nicht-schweizer Sprachen beschreiben. Die Länder (bzw. Landescodes), in denen irgendeine nicht-schweizer Sprache gesprochen wird, bekommt man also folgendermassen:

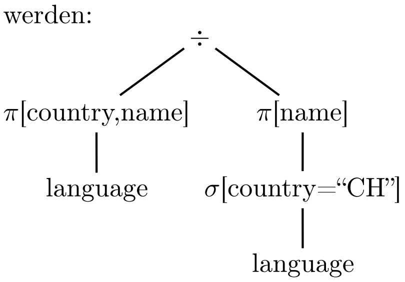


Bleibt nun, die Komplementmenge, eben die Länder in denen nur schweizer Sprachen gesprochen werden, zu bestimmen:



... schönes Beispiel dafür, wie man Subqueries in der WHERE-Klausel in der Algebra durch Joins realisiert.

d) "alle" – was wieder mit Division gelöst wird. $\pi[country, name](language)$ durch die Menge der schweizer Sprachen dividieren. Übrig bleiben die Länder die mit allen diesen Sprachen genannt werden:



Aufgabe 8 (Äquivalenzen: Join, Division, Differenz) Seien $R(\bar{X}), S(\bar{Y})$ Relations-Schemata. Zeigen oder widerlegen Sie:

(a) Sei $\bar{X} \cap \bar{Y} = \emptyset$.

$$(R \bowtie S) \div S \equiv R.$$

(b) Sei $\bar{X} = \bar{Y}$ und $\bar{Z} \subseteq \bar{X}$.

$$\pi[\bar{Z}](R - S) \equiv \pi[\bar{Z}]R - \pi[\bar{Z}]S.$$

(zu a) Nach Voraussetzung sind $R(\bar{X})$ und $S(\bar{Y})$ Relationen, die kein gemeinsames Attribut haben. Damit gilt $(\bar{X} \cup \bar{Y}) - \bar{Y} = \bar{X}$.

Behauptung: dann gilt: $(R \bowtie S) \div S = R$

Definition der Division: Sei $\bar{Y} \subset \bar{X}$, $\bar{Z} = \bar{X} - \bar{Y}$.

$r \div s = \{\mu \in \mathsf{Tup}(\bar{Z}) \mid \{\mu\} \times s \subseteq r\} = \pi[\bar{Z}](r) - \pi[\bar{Z}]((\pi[\bar{Z}](r) \times s) - r)$.

Man kann entweder die Umschreibung der Division in Algebra-Operatoren, oder die Definition verwenden:

$$\begin{array}{ll}
(R \bowtie S) \div S & \text{keine gemeinsamen Attribute} \\
= (R \times S) \div S & \text{Umschreibung durch Algebra-Operatoren} \\
= \pi[\bar{X}](R \times S) - \pi[\bar{X}]((\pi[\bar{X}](R \times S) \times S) - R \times S) & \\
= R - \pi[\bar{X}]((R \times S) - (R \times S)) & \\
= R - \pi[\bar{X}](\emptyset) & \\
= R &
\end{array}$$

(dies ist die interessantere, aber längere Beweis), oder

$$\begin{array}{ll}
(R \bowtie S) \div S & \text{keine gemeinsamen Attribute} \\
= (R \times S) \div S & \text{Definition der Division: mit } R \times S \subseteq \mathsf{Tup}(\overline{XY}) \\
= \{\mu \in \mathsf{Tup}(\bar{X}) \mid \mu \times S \subseteq (R \times S)\} & \\
= \{\mu \in \mathsf{Tup}(\bar{X}) \mid \mu \in R\} = R &
\end{array}$$

(zu b) $\pi[\bar{Z}](R - S) \neq \pi[\bar{Z}](R) - \pi[\bar{Z}](S)$

Bsp:

Sei $\bar{X} = \bar{Y} = \{A, B\}$ und $\bar{Z} \subseteq \bar{X} = \{A\}$

R	
A	B
1	2

S	
A	B
1	3

$\pi[A](R - S) = \pi[A](R) = \{(A : 1)\}$ (da $R - S = R$ ist), und $\pi[A](R) - \pi[A](S) = \{(A : 1)\} - \{(A : 1)\} = \emptyset$

Aufgabe 9 (Algebra: Minimale- und Maximale Anzahl von Tupeln) Die Relationen $R(\bar{X})$ und $S(\bar{Y})$ enthalten n bzw. m Tupel. Wie groß ist die maximale und minimale Anzahl von Tupeln, die das Ergebnis folgender Operationen (bei geeigneten $\bar{X}, \bar{Y}$) enthalten kann?

a) $R \cup S$

b) $R \bowtie S$

c) $\sigma[C](R) \times S$, für eine Bedingung C

d) $\pi[\bar{Y}](R) - S$

e) $R \div S$

---

a) Nur zulässig wenn $\bar{X} = \bar{Y}$.

max: $n + m$ (keine gemeinsamen Tupel)

min: $max\{n, m\}$ (Teilmengenbeziehung)

b) max: $n * m$, wenn entweder $\bar{X} \cap \bar{Y} = \emptyset$ oder im natürlichen Join jedes Tupel R mit jedem Tupel aus S auf den gemeinsamen Attributen $\bar{X} \cap \bar{Y}$ zusammenpasst (d.h., diese Attribute haben in allen Tupeln beider Relationen den selben Wert).

min: 0 (natürliches Join mit $\bar{X} \cap \bar{Y} \supsetneq \emptyset$, bei dem aber nichts übrigbleibt, da die Einträge der gemeinsamen Spalte(n) nicht zusammenpassen.

Hinweis: man sieht, dass ein Join also die Ergebnismenge sowohl deutlich vergrößern, als auch einschränkend wirken kann.

- Verhältnis zwischen $|R \bowtie S|$ und $|R|$ bezeichnet man als *Selektivität* des Joins (bzgl. R).
- ist $\bar{X} \cap \bar{Y}$ z.B. Schlüssel von S und Fremdschlüssel in R, so hat man $|R \bowtie S| = |R|$.

c) max: $n * m$ (alle Tupel in R erfüllen C)
   min: 0 (kein Tupel in R erfüllt C)

d) Nur sinnvoll wenn $\bar{Y} \subseteq \bar{X}$.
   max: n, wenn $S = \emptyset$, oder zumindest $S \cap \pi[\bar{Y}](R) = \emptyset$, und bei der Projektion keine Duplikate (die wegfallen würden) entstehen.
   min: 0, wenn $\pi[\bar{Y}](R) \subseteq S$.

e) max: n/m – ganzzahlige Division, wenn alle Werte $\pi[\bar{X} \setminus \bar{Y}](R)$ mit jedem der Werte aus S in R vorkommen.
   Insbesondere, wenn $|S| = m = 1$ ist, und jedes Tupel aus R mit diesem Wert zusammen vorkommt, gilt $R \div S = \pi[\bar{X} \setminus \bar{Y}](R)$ und $|R \div S| = n$.
   Wenn noch extremer $|S| = \emptyset$ ist, ist $R \div S = \pi[\bar{X} \setminus \bar{Y}](R)$, und ebenfalls $|R \div S| = n$.
   min: 0, wenn kein Wert die durch die Division bzgl. S gestellte "für alle"-Bedingung erfüllt.